

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-302267

(43)Date of publication of application : 14.11.1995

(51)Int.Cl.

G06F 17/30

(21)Application number : 06-295053

(71)Applicant : NIPPON TELEGR & TELEPH CORP
<NTT>

(22)Date of filing : 29.11.1994

(72)Inventor : NAGAYOSHI TAKESHI

(30)Priority

Priority number : 06 38757

Priority date : 09.03.1994

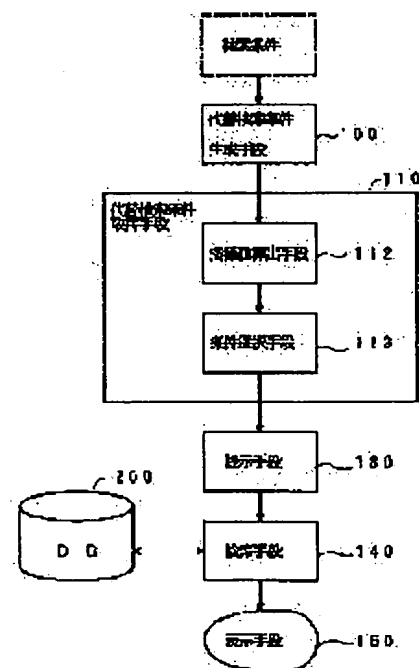
Priority country : JP

(54) INFORMATION RETRIEVAL DEVICE, INFORMATION RETRIEVAL METHOD AND KNOWLEDGE ACQUISITION SYSTEM

(57)Abstract:

PURPOSE: To provide an information retrieval device, an information retrieval method and a knowledge acquisition system capable of reducing the troublesomeness of operations required for information retrieval and improving the success rate of the information retrieval.

CONSTITUTION: An alternate retrieval condition acquisition means 110 for obtaining alternate retrieval conditions for maximizing an index value for indicating that a retrieval solution set is provided with more data requested by a user and less data not required by the user at the time of retrieving a data base 200 with the alternate retrieval conditions to be candidates and a retrieval means 140 for retrieving the data base 200 again with the alternate retrieval conditions obtained by the alternate retrieval condition acquisition means 110 are provided.



LEGAL STATUS

[Date of request for examination] 11.11.1999

[Date of sending the examiner's decision of rejection] 22.04.2003

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3457405

[Date of registration] 01.08.2003

[Number of appeal against examiner's decision of rejection] 2003-09079

[Date of requesting appeal against examiner's decision of rejection] 22.05.2003

Submitted in 09/552,589

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

| (51) Int.Cl. ⁶ | 識別記号 | 庁内整理番号 | F I | 技術表示箇所 |
|---------------------------|------|---------|-----------------|---------|
| G 0 6 F 17/30 | | 9194-5L | G 0 6 F 15/ 403 | 3 2 0 C |

審査請求 未請求 請求項の数19 O L (全 20 頁)

| | | | |
|--------------|------------------|----------|--|
| (21) 出願番号 | 特願平6-295053 | (71) 出願人 | 000004226 日本電信電話株式会社 東京都千代田区内幸町一丁目1番6号 |
| (22) 出願日 | 平成6年(1994)11月29日 | (72) 発明者 | 永吉 剛 東京都千代田区内幸町1丁目1番6号 日 本電信電話株式会社内 |
| (31) 優先権主張番号 | 特願平6-38757 | (74) 代理人 | 弁理士 伊東 忠彦 |
| (32) 優先日 | 平6(1994)3月9日 | | |
| (33) 優先権主張国 | 日本 (J P) | | |

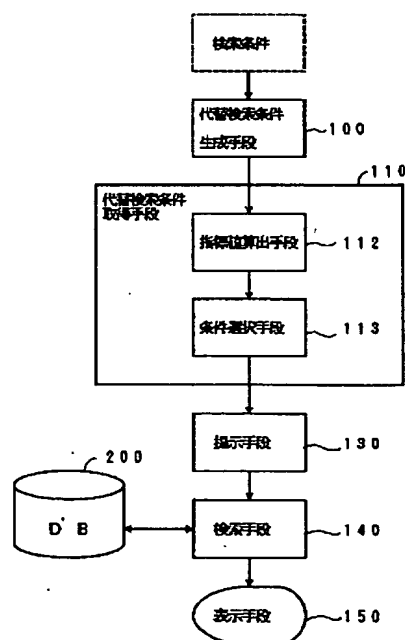
(54) 【発明の名称】 情報検索装置、情報検索方法及び知識獲得システム

(57) 【要約】

【目的】 本発明の目的は、情報検索に要する操作手数を軽減し、かつ情報検索の成功率を高めることが可能な情報検索装置、情報検索方法及び知識獲得システムを提供することである。

【構成】 本発明は、候補となる代替検索条件でデータベース200を検索したとき、検索解集合に利用者が要求するデータがより多く、且つ利用者が要求しないデータがより少なく含まれることを指す指標値が最大となる代替検索条件を求める代替検索条件取得手段110と、代替検索条件取得手段110によって得られた代替検索条件でデータベース200を再検索する検索手段140を有する。

本発明の情報検索装置の処理構成図



【特許請求の範囲】

【請求項1】 複数の項目からなる検索条件を満たすデータをデータベースから検索する場合に、該データベース中に該検索条件を満たすデータが存在しない場合には、該検索条件を緩和した代替検索条件を複数種類生成し、該代替検索条件でデータベースを検索する情報検索装置において、

候補となる代替検索条件で前記データベースを検索したとき、検索解集合に利用者が要求するデータがより多く、且つ該利用者が要求しないデータがより少なく含まれることを指す指標値が最大となる代替検索条件を求める代替検索条件取得手段と、
前記代替検索条件取得手段により前記データベースを再検索する検索手段を有することを特徴とする情報検索装置。

【請求項2】 前記代替検索条件取得手段は、ある代替検索条件を満たすデータが実際に前記データベース内に存在する確率 P の、前記利用者が要求するデータが前記データベース内に存在しないと仮定した場合に、該代替検索条件を満たすデータが前記データベース内に存在する確率の推定値 (P') に対する比を表す指標値 (P/P') が最大となる代替検索条件を求める請求項1記載の情報検索装置。

【請求項3】 前記代替検索条件取得手段は、ある代替検索条件を満たすデータが実際に前記データベース内に存在する確率 P の利用者が要求するデータが前記データベース内に存在しないと仮定した場合に、該代替検索条件を満たすデータが前記データベース内に存在する確率の推定値 (P') に対する比を表す指標値 (P/P') を算出する指標値算出手段と、
一件以上のデータベースの検索解が存在する全ての代替検索条件について指標値を算出し、該推定値 P' に対して該確率 P が最大となる代替検索条件を選択する条件選択手段と、
該条件選択手段により選択された代替検索条件を再検索の検索条件として提示する提示手段とを有する請求項2記載の情報検索装置。

【請求項4】 前記情報検索装置の前記条件選択手段で選択された代替検索条件で再検索を行った検索結果を表示する表示手段を有する請求項3記載の情報検索装置。

【請求項5】 前記情報検索装置において、
複数項目の検索キーを有するデータを格納したデータベースから各項目の検索キーの論理積からなる検索条件を満たすデータを検索する場合、
前記指標値算出手段は、
代替検索条件 Q ($Q = A \cap B \cap C \cap \dots$) を満たすデータが存在する確率を P とし、該代替検索条件 Q の各検索キー A , B , $C \dots$ のみを満たすデータが存在する確率を P_A , P_B , P_C , \dots とし、
前記代替検索条件 Q の評価関数 $F(Q)$ を、

【数1】

$$F(Q) = \frac{P}{P_A \times P_B \times P_C \times \dots}$$

とし、

任意の代替検索条件 Q_1 , Q_2 に対して $F(Q_1) > F(Q_2)$ ならば $S(Q_1) > S(Q_2)$ が成り立つ任意の評価関数 $S(Q)$ を、前記代替検索条件 Q の指標値として算出する請求項3記載の情報検索装置。

【請求項6】 複数の項目からなる検索条件を満たすデータをデータベースから検索する場合に、該データベース中に該検索条件を満たすデータが存在しない場合には、該検索条件を緩和した代替検索条件を複数種類生成し、該代替検索条件でデータベースを検索する情報検索システムにおいて、

候補となる代替検索条件で該データベースを検索したとき、検索解集合に利用者が要求するデータがより多く、且つ利用者が要求しないデータがより少なく含まれることを指す指標値が最大となる代替検索条件を求めるステップと、

該代替検索条件で該データベースを再検索するステップからなることを特徴とする情報検索方法。

【請求項7】 前記情報検索方法は、ある代替検索条件を満たすデータが実際に前記データベース内に存在する確率 (P) の、利用者が要求するデータが前記データベース内に存在しないと仮定した場合に該代替検索条件を満たすデータが前記データベース内に存在する確率の推定値 (P') に対する比を表す指標値 (P/P') が最大となる代替検索条件を求める請求項6記載の情報検索方法。

【請求項8】 前記情報検索方法は、指標値が最大となる代替検索条件を再検索の検索条件として提示する提示ステップを更に有する請求項6記載の情報検索方法。

【請求項9】 複数の項目からなる検索条件を満たすデータをデータベースから検索する情報検索システムにおいて、

該データベースに該検索条件を満たすデータが存在しない場合に、該検索条件を修正した代替検索条件を複数種類生成する代替検索条件生成ステップと、

複数項目の検索キーを有するデータを格納したデータベースから各項目の検索キーの論理積からなる検索条件を満たすデータを検索する検索ステップと、

ある代替検索条件を満たすデータが実際に前記データベース内に存在する確率 (P) の、利用者が要求するデータが前記データベース内に存在しないと仮定した場合に前記代替条件を満たすデータが前記データベース内に存在する確率の推定値 (P') に対する比を表す指標値 (P/P') を算出する算出ステップと、

1件以上のデータベースの検索解が存在する全ての代替検索条件について指標値を算出し、該推定値 P' に対し

て該確率Pが最大となる代替検索条件を選択する最適条件判断ステップと、
再検索の検索条件として提示する提示ステップよりなる請求項7記載の情報検索方法。

【請求項10】 前記代替検索条件生成ステップは、前記データベースに検索条件を満たすデータが存在しない場合に、該検索条件を部分的に緩和して代替検索条件を生成する請求項9記載の情報検索方法。

【請求項11】 前記情報検索システムにおいて、複数項目の検索キーを有するデータを格納したデータベースから、各項目の検索キーの論理積からなる検索条件を満たすデータを検索する場合、
前記算出ステップは、
代替検索条件Q ($Q = A \cap B \cap C \cap \dots$) を満たすデータが存在する確率をPとし、Qの各検索キーA, B, C...のみを満たすデータが存在する確率を P_A , P_B , P_C とし、
代替検索条件Qの評価関数F (Q) を

【数2】

$$F(Q) = \frac{P}{P_A \times P_B \times P_C \times \dots}$$

とし、

任意の代替検索条件 Q_1 , Q_2 に対して $F(Q_1) > F(Q_2)$ ならば、 $S(Q_1) > S(Q_2)$ が成り立つ任意の評価関数S (Q) を、代替検索条件Qの指標値として算出する請求項9記載の情報検索方法。

【請求項12】 多数の利用者が利用する情報検索サービスにおける検索キーの別名 (エイリアス) を取得する知識獲得システムにおいて、

請求項1乃至11記載の情報検索装置と、
前記検索キーの別名として登録すべきデータと該データの適切な別名を判断する判定手段とを有することを特徴とする知識獲得システム。

【請求項13】 前記判定手段は、
前記情報検索装置によって入力条件を自動修正し、優先的に出力されたデータが当初の要求に適合すると利用者が承認したことを確認する適合確認手段と、
前記入力条件と適合性が確認されたデータの対応関係の有効性を評価する有効性評価手段と、
前記信頼性が高いと評価された場合には、前記データの検索キーの別名として前記入力条件を登録する別名登録手段を有する請求項12記載の知識獲得システム。

【請求項14】 前記判定手段は、
修正された入力条件と適合性が確認されたデータの対応関係の累積頻度と、該データが検索された累積総数を記憶する累積頻度記憶手段を有する請求項12記載の知識獲得システム。

【請求項15】 前記有効性評価手段は、
前記累積頻度記憶手段の累積頻度により前記情報検索装置で検索された前記検索結果の対応関係の有効性を評価

する請求項12及び13記載の知識獲得システム。

【請求項16】 多数の利用者が利用する情報検索サービスにおける検索キーの別名 (エイリアス) を取得する知識取得システムにおいて、

請求項1乃至11記載の情報検索装置と、
実際に問い合わせが生じた入力条件のみを別名として登録する別名登録手段を有することを特徴とする知識獲得システム。

【請求項17】 前記別名登録手段は、
前記情報検索装置によって入力条件を自動修正し、優先的に出力されたデータが当初の要求に適合すると利用者が承認したことを確認する適合確認手段と、
前記入力条件と適合性が確認されたデータの対応関係の有効性を評価する有効性評価手段と、
前記有効性評価手段により前記信頼性が高いと評価された場合には、前記データの検索キーの別名として前記入力条件を登録する別名登録手段とを有する請求項16記載の知識獲得システム。

【請求項18】 前記別名登録手段は、
修正された入力条件と適合性が確認されたデータの対応関係の累積頻度と、該データが検索された累積総数を記憶する累積頻度記憶手段を有する請求項16記載の知識獲得システム。

【請求項19】 前記有効性評価手段は、
前記累積頻度記憶手段の累積頻度により前記情報検索装置で検索された前記検索結果の対応関係の有効性を評価する請求項16及び17記載の知識獲得システム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、情報検索装置、情報検索方法及び知識獲得システムに係り、特に、曖昧な情報を含む検索条件から利用者の要求するデータを自動検索する情報検索装置、情報検索方法及び知識獲得システムに関する。

【0002】

【従来の技術】 従来、情報検索システムにおいて、データベース内に入力した検索条件を全て満たすデータが存在しない場合は、入力した検索条件に修正を加えてデータベースの再検索が行われる。多くの情報検索システムでは、利用者が自発的に検索条件を修正し、再検索を行っている。

【0003】 また、検索条件の修正を支援する手法として、特定のルールに基づき当初の検索条件を緩和した検索条件で再検索を行う自動緩和方式が用いられている。以下に、従来の第1のシステムとして、電話帳検索システムを例に説明する。全ての掲載データが名義文字列、市区と町について階層的に付与された住所、及び職業の3項目の検索キーを持つ。名義については前方一致検索を行い、入力された全ての条件を満たす掲載データのみが検索される。

【0004】以下の説明において、ある検索条件を満たすデータの集合を「検索解」と呼ぶ。例えば、検索条件として、

入力条件1

「住所：千代田区 北の丸公園

職業：美術館

名義：キンダイビジュツカン」

が入力されたものとする。以下、入力された検索条件を「入力条件」と呼ぶ。

【0005】入力条件は、特定の掲載データを検索解として得ることを要求するものであり、入力条件の要求を満たすデータを以下「適合解」と呼ぶ。また、適合解以外の検索解を以下「不適合解」と呼ぶ。上記入力条件の適合解の検索キーは次のものとする。

掲載条件1

「住所：千代田区 北の丸公園

職業：美術館

名義：コクリツキンダイビジュツカン」

以下、適合解の検索キーを「掲載条件」と呼ぶ。適合解を検索できなかった場合、検索条件の指定範囲を修正する手法として、以下のような操作がある。

「キンダイビジュツカン」→「キンダイビジュツ」、

「キンダイ」、…

「千代田区 北の丸公園」→「千代田区」、「（入力なし）」、…

「美術館」→「（入力なし）」

さらに、各項目について上記のような緩和方法を組み合わせることによって、複数種類の検索条件の生成が可能となる。以下、修正された検索条件を「代替検索条件」と呼ぶ。上記の例のような入力条件の部分的緩和によって生成され得る代替検索条件を図13に全て列挙する。

【0006】最適な検索結果は適合解の数が最大で、不適合解の数が最小の検索解である。適合解の数がより多く不適合解の数がより少ない検索結果を与える検索条件を、以下「適切」な検索条件と呼ぶ。入力条件から掲載条件に含まれない条件を削除した代替検索条件の検索解には適合解が含まれる。しかし、適合解が検索解に含まれていても、条件を削除するに従い、不適合解の数が増加するため、最小限の条件削除が望ましい。即ち、上記複数種類の代替検索条件のうち、最も適切な代替検索条件は、掲載条件に含まれない条件のみを削除した代替検索条件である。図13に列挙した中で、上記入力条件1に対する最適な代替検索条件は、以下の代替検索条件である。

【0007】代替検索条件1

「住所：千代田区 北の丸公園

職業：美術館

名義：（入力なし）」

従来の条件自動修正方式は、特定のルールに基づき、上記代替検索条件のいずれかを自動的に生成し、再検索を

行う方法である。例えば、以下のような自動緩和ルールがある。

【0008】ルール1：「入力条件の名義が6文字以上であり、検索解が存在しない場合は、名義を先頭4文字に末尾切捨てとする」

ルール2：「入力条件の住所が“町”までの指定であり、検索解が存在しない場合は住所の“町”を切捨てる」

ルール3：「ルール2の修正による代替検索条件でも検索解が存在しない場合はルール1の修正を適用する」

このようなルールは、名義の末尾部分や住所の詳細部分には、曖昧さが生じやすいという経験的規則に基づき採用されている。例えば、以下の掲載条件2に対して入力条件2が入力された場合、上記ルール3を適用し、代替検索条件2を生成し、再検索することによって要求する掲載データを検索できる。

【0009】入力条件2

「住所：港区 三田

職業：出版社

名義：エヌティティーシュッパン」

掲載条件2

「住所：港区 芝

職業：出版社

名義：エヌティティシュッパン」

代替検索条件2

「住所：港区

職業：出版社

名義：エヌティ」

次に、従来の第2のシステムを説明する。

【0010】企業名の略称、通称、あるいは意味的に類似した職業名のように、登録されている検索キーとは異なる条件を、多くの利用者が入力する掲載データがある。例えば、以下の掲載条件3あるいは4に対して、入力条件3或いは4では適合解を検索できない。

掲載条件3

名義：ニホンデンシンデンワ

入力条件3

名義：エヌティティ

掲載条件4

職業：レストラン

入力条件4

職業：フランス料理店

このような場合、掲載データの検索キーには、頻繁に入力される入力条件を、別名（エイリアス）として以下のように共に登録する対策がなされている。

掲載条件3'

名義：ニホンデンシンデンワ、エヌティティ

掲載上演4'

職業：レストラン、フランス料理店

従来、別名の登録に際して、「どの掲載条件にどのよう

な別名を登録するか」についての知識の獲得は、多くの場合データベースの運用を通して人手によって経験的になされている。

【0011】或いは、複数の単語によって構成される企業名については省略されやすい単語の位置に規則性があることを利用して企業名を「地名」、「職業名」、「固有名」その他の“意味”を付与した単語単位に分割する解析処理を行い、解析結果の意味配列パターンに応じてどの単語が省略されやすいかを指定する知識テーブルを参照することによって別名を自動生成する方式が提案されている。例えば、

掲載条件5

名義：ハチオウジンノウキョウ（八王子市農協）を解析すると、意味配列パターンは「地名+地名語尾+職業名」であり、この意味配列パターンに対応する別名は「第1単語+第3単語」と登録されていれば、

掲載条件5'

名義：ハチオウジンノウキョウ（八王子市農協）、ハチオウジノウキョウ（八王子農協）を生成することができる。

【0012】

【発明が解決しようとする課題】しかしながら、上記従来の方法では、入力条件に応じて入力条件の適切な修正方法を判断できないため、固定された修正ルールの予想に反する曖昧さが問い合わせに生じた場合は、不適切な代替検索条件を生成し、適合解を含まない検索結果や、多数の不適合解を出力するという問題がある。

【0013】例えば、上記入力条件1が入力された場合、上記ルール1～3のいずれを用いても適合解を検索できない。入力条件1に対して最適な代替検索条件1を生成するには、以下のルール4が必要である。

ルール4：「検索解が存在しない場合は、名義を全て削除する」

しかし、ルール4を入力条件2に適用すると再び不適合解のみを出力する。一方、

ルール5：「検索解が存在しない場合は、名義と職業を全て削除する」

上記ルール5のような場合は多くの条件を削除する方法であり、適合解が検索解に含まれる可能性が高いが、不適合解も多数含まれるため、不適である。従って、結局、利用者自らが条件を修正しなければならないという問題がある。

【0014】このような情報検索システムの操作は、検索経験の少ない利用者にとって困難であり、著しく利便性を欠く。その結果、利用者は検索を断念するか、または再検索のために多くの時間を費やすことになる。さらに、従来の第2のシステムについては、人手による別名の登録は、主観や偶然に依存するため登録基準が不統一となること、及びデータベースが大規模な場合には、経済的効率が悪いこと等が問題となる。

【0015】さらに、意味配列パターンの知識ベースに適合する企業名を全て別名登録すれば、実際には問い合わせを受けたくないような別名が登録されるケースが発生し得るという問題がある。本発明は、上記の点に鑑みてなされたもので、上記従来の問題点を解決し、情報検索に要する操作手数を軽減し、かつ情報検索の成功率を高めることが可能な情報検索装置、情報検索方法及び知識獲得システムを提供することを目的とする。

【0016】更なる本発明の目的は、入力された条件を満たすデータがデータベース中に存在しない場合、入力条件を緩和することによって複数発生する代替検索条件の中から最適な代替検索条件を推定し、利用者の要求するデータを出力することができる情報検索装置、情報検索方法及び知識獲得システムを提供することである。本発明の更なる目的は、検索キーに対応する別名を自動的に登録可能な知識獲得システムを提供することである。

【0017】

【課題を解決するための手段】図1は、本発明の情報検索装置及び情報検索方法の原理構成図である。本発明は、入力条件を満たすデータをデータベース200から検索する場合に、データベース200中に検索条件を満たすデータが存在しない場合には、検索条件を修正した代替検索条件を複数種類生成し、代替検索条件データベース200を検索する情報検索装置において、候補となる代替検索条件でデータベース200を検索したとき、検索解集合に利用者が要求するデータがより多く、且つ、利用者が要求しないデータがより少なく含まれることを指す指標値が最大となる代替検索条件を求める代替検索条件取得手段110と、代替検索条件取得手段110によって得られた代替検索条件でデータベース200を再検索する検索手段140を有する。

【0018】本発明は、情報検索装置の代替検索条件取得手段110は、ある代替検索条件を満たすデータが実際にデータベース内に存在する確率（P）の、利用者が要求するデータがデータベース内に存在しないと仮定した場合に代替検索条件を満たすデータがデータベース内に存在する確率の推定値（P'）に対する比を表す指標値（ P/P' ）が最大となる代替検索条件を求める。

【0019】また、本発明は、入力条件を満たすデータをデータベース200から検索する検索手段140と、データベース200中に入力条件を満たすデータが存在しない場合には、検索条件を修正した代替検索条件を複数種類生成する代替検索条件生成手段100を有する情報検索装置において、候補となる代替検索条件でデータベース200を検索したとき、ある代替検索条件を満たすデータが実際にデータベース内に存在する確率（P）の、利用者が要求するデータがデータベース内に存在しないと仮定した場合に代替検索条件を満たすデータがデータベース内に存在する確率の推定値（P'）に対する比を表す指標値（ P/P' ）を算出する指標値算出手段

1 1 2と、指標値が最大となる代替検索条件を選択する条件選択手段1 1 3と、条件選択手段1 1 3で選択された検索条件で検索手段1 4 0により再検索を行った検索結果を表示する表示手段1 5 0とを有する。

【0 0 2 0】また、本発明は、条件選択手段1 1 3で選択された再検索の検索条件を提示する提示手段1 3 0を有する。また、複数項目の検索キーを有するデータを格納したデータベースから各項目の検索キーの論理積からなる検索条件を満たすデータを検索する場合、指標値算出手段1 1 2は、代替検索条件 Q ($Q = A \cap B \cap C \cap \dots$) を満たすデータが存在する確率を P とし、 Q の各検索キー A , B , $C \dots$ のみを満たすデータが存在する確率を P_A , P_B , P_C とし、代替検索条件 Q の評価関数 F (Q) を

【0 0 2 1】

【数3】

$$F(Q) = \frac{P}{P_A \times P_B \times P_C \times \dots}$$

【0 0 2 2】とし、任意の代替検索条件 Q_1 , Q_2 に対して $F(Q_1) > F(Q_2)$ ならば、 $S(Q_1) > S(Q_2)$ が成り立つ任意の評価関数 $S(Q)$ を、代替検索条件 Q の指標値として算出する。本発明の情報検索方法は、入力条件を満たすデータをデータベース2 0 0から検索する場合に、データベース2 0 0中に入力条件を満たすデータが存在しない場合には、検索条件を修正した代替検索条件を複数種類生成し、代替検索条件でデータベース2 0 0を検索する情報検索システム2 0 0において、候補となる代替検索条件でデータベース2 0 0を検索したとき、検索解集合に利用者が要求するデータがより多く、且つ利用者が要求しないデータがより少なく含まれることを指す指標値が最大となる代替検索条件を求め、代替検索条件でデータベース2 0 0を再検索する。

【0 0 2 3】図2は、本発明の情報検索方法の原理を説明するためのフローチャートである。また、本発明の情報検索方法は、ある代替検索条件を満たすデータが実際にデータベース内に存在する確率 (P) の、利用者が要求するデータがデータベース内に存在しないと仮定した場合に、代替検索条件を満たすデータがデータベース内に存在する確率の推定値 (P') に対する比を表す指標値 (P/P') が最大となる代替検索条件でデータベース2 0 0を再検索する。

【0 0 2 4】また、本発明の情報検索方法は、入力条件を満たすデータをデータベース2 0 0から検索する情報検索システムにおいて、データベース2 0 0中に入力条件を満たすデータが存在しない場合に、検索条件を修正した代替検索条件を複数種類生成し (ステップ1)、代替検索条件でデータベース2 0 0を再検索し (ステップ2)、ある代替検索条件を満たすデータが実際にデータベース内に存在する確率 (P) の、利用者が要求するデ

ータがデータベース内に存在しないと仮定した場合に、代替検索条件を満たすデータがデータベース内に存在する確率の推定値 (P') に対する比を表す指標値 (P/P') を複数の代替検索条件について算出し (ステップ3)、指標値が最大となる代替検索条件を選択し (ステップ4)、選択された検索条件で再検索を行った検索結果を表示する (ステップ5)。

【0 0 2 5】また、複数項目の検索キーを有するデータを格納したデータベースから、各項目の検索キーの論理積からなる検索条件を満たすデータを検索する場合、本発明の指標値算出ステップ (ステップ3) は、代替検索条件 Q ($Q = A \cap B \cap C \cap \dots$) を満たすデータが存在する確率を P とし、 Q の各検索キー A , B , $C \dots$ のみを満たすデータが存在する確率を P_A , P_B , P_C とし、代替検索条件 Q の評価関数 $F(Q)$ を

【0 0 2 6】

【数4】

$$F(Q) = \frac{P}{P_A \times P_B \times P_C \times \dots}$$

【0 0 2 7】とし、任意の代替検索条件 Q_1 , Q_2 に対して $F(Q_1) > F(Q_2)$ ならば、 $S(Q_1) > S(Q_2)$ が成り立つ任意の評価関数 $S(Q)$ を代替検索条件 Q の指標値として算出する。図3は、本発明の知識獲得システムの原理構成図である。

【0 0 2 8】また、本発明の知識獲得システムは、多数の利用者が利用する情報検索サービスにおける検索キーの別名 (エイリアス) を取得する知識獲得システムにおいて、上記の情報検索装置3 0 0と、情報検索装置3 0 0によって入力条件を自動修正し、優先的に出力されたデータが当初の要求に適合すると利用者が承認したことを確認する適合確認手段3 1 0と、入力条件と適合性が確認されたデータの対応関係の有効性を評価する有効性評価手段3 2 0と、信頼性が高いと評価された場合には、データの検索キーの別名として入力条件を登録する別名登録手段3 3 0を有し、別名を登録すべきデータとデータの適切な別名を自動的に判断する。

【0 0 2 9】また、上記の適合確認手段は、修正された入力条件と適合性が確認されたデータの対応関係の累積頻度と、データが検索された累積総数を記憶する累積頻度記憶手段3 4 0を有する。また、上記の有効性評価手段3 2 0は、累積頻度記憶手段3 4 0の累積頻度により検索結果の有効性を評価する。

【0 0 3 0】また、本発明の知識獲得システムは、数の利用者が利用する情報検索サービスにおける検索キーの別名 (エイリアス) を取得する知識取得システムにおいて、情報検索装置によって入力条件を自動修正し、優先的に出力されたデータが当初の要求に適合すると利用者が承認したことを確認する適合確認手段と、修正された入力条件と適合性が確認されたデータの対応関係の累積頻度と、データが検索された累積総数を記憶する累積頻

度記憶手段と、入力条件と検索結果の対応関係の有効性を評価する有効性評価手段と、有効性評価手段により信頼性が高いと評価された場合には、データの検索キーの別名として入力条件を登録する別名登録手段とを有し、実際に問い合わせが生じた入力条件のみを別名として登録する別名登録手段330を有する。

【0031】

【作用】上記のように、請求項1から請求項11の発明は、最も指標値がS(Q)が大きい代替検索条件を選択し、このように選択された代替検索条件は、検索解集合に利用者が要求するデータがより多く、かつ利用者が要求しないデータがより少なく含まれる代替検索条件である。従って、最適な代替検索条件による再検索の結果を出力することができる。

【0032】また、請求項3及び8の発明は、選択された代替検索条件で再検索を行う前に、代替検索条件を提示し、その有効性を利用者に確認することによってさらに、条件修正の有効性を向上すると共に、不要なデータベース検索の回数を減少させることが可能となる。また、請求項12の発明は、上記の情報検索装置または情報検索方法を用いた知識獲得システムであり、別名を登録すべきデータとデータの適切な別名を自動的に判断することにより、登録基準を統一することが可能となる。

【0033】さらに、請求項16の発明は、上記の情報検索装置または情報検索方法を用いた知識獲得システムであり、実際に問い合わせが生じた入力条件のみを別名として登録するため、実際に問い合わせを受けないような別名が登録されるケースが発生しない。

【0034】

【実施例】以下、図面と共に本発明の実施例を詳細に説明する。代替検索条件の適切さは、本発明の情報検索装置及び情報検索方法における指標値によって表されることを以下に説明する。データベース内のデータの素数をN、データベース中に含まれる適合解の数をXとする。また、代替検索条件Qを満たす検索解の数をz、適合解の数をx、不適合解の数をyとする。このとき、x、y、zには、以下の関係がある。

【0035】 $z = x + y$

代替検索条件の適切さとは、xがより多く、yがより少ないことを指すが、x、yを直接知ることはいできない。但し、検索解の数zは既知である。このとき、本発明の情報検索装置及び情報検索方法における「代替検索条件Qを満たすデータが実際にデータベース内に存在する確率」Pは、

【0036】

【数5】

$$P = \frac{z}{N}$$

【0037】によって求められる。また、適合解が存在する確率 P_x は、

【0038】

【数6】

$$P_x = \frac{x}{N}$$

【0039】不適合解が存在する確率 P_y は、

【0040】

【数7】

$$P_y = \frac{y}{N}$$

10 【0041】である。従って、P、 P_x 、 P_y の間には、

$$P = P_x + P_y$$

の関係がある。本発明の情報検索装置及び情報検索方法における「利用者が要求するデータがデータベース内に存在しないと仮定した場合に、代替検索条件を満たすデータがデータベース内に存在する確率の推定値」 P' は、データベースから適合解を取り除いたとき、残る $N' = N - X$ 件のデータの中に代替検索条件Qを満たす検索解が存在する確率の推定値である。

20 【0042】電話帳のように $N \gg X$ が成り立つ場合、 P' は適合解を含まないN件のデータベース中に代替検索条件Qを満たす検索解が存在する確率、即ち、 P_y に略等しく、

$$P' \approx P_y$$

が成り立つ。

【0043】本発明の情報検索装置及び情報検索方法における代替検索条件Qの指標値をS(Q)とすると、S(Q)はPの P' に対する比であり、S(Q)を P_x と P_y を用いて表せば、以下ようになる。

30 【0044】

【数8】

$$S(Q) = \frac{P}{P'} \approx \frac{P_x + P_y}{P_y} = 1 + \frac{P_x}{P_y}$$

【0045】ここで、2つの代替検索条件 Q_1 、 Q_2 があり、

$$P_x(Q_2) / P_x(Q_1) = s \cdots (0 \leq s \leq 1)、$$

$$P_y(Q_2) / P_y(Q_1) = t \cdots (t \geq 1)$$

40 の関係があるとする。即ち、 Q_2 は、 Q_1 と比較して適合解の数が同数または、少なく、不適合解の数が同数または多い検索結果を与えるため、 Q_1 が Q_2 より適切な代替検索条件であるものとする。このとき、

【0046】

【数9】

$$S(Q_1) = 1 + \frac{P_x(Q_1)}{P_y(Q_1)},$$

$$S(Q_2) = 1 + \frac{s \times P_x(Q_1)}{t \times P_y(Q_1)}$$

【0047】よって、sとtの範囲に関する上記の定義より、

$$S(Q_1) \geq S(Q_2)$$

が成り立つ。従って、指標値S(Q)を代替検索条件の適切さ、即ち、検索解集合に利用者が要求するデータがより多く、且つ利用者が要求しないデータがより少なく含まれることの指標値として用いることができる。

【0048】次に、上記指標値S(Q)が算出可能であることを、上記電話帳データベースの事例のように代替検索条件Qが住所A、職業B、名義Cの論理積 $Q = A \cap B \cap C$ からなる場合について説明する。まず、P'は、以下のように推定できる。A、B、Cのみを満たすデータの数をそれぞれa、b、cとすると、A、B、Cを満たすデータが存在する確率 P_A 、 P_B 、 P_C はそれぞれ、

【0049】

【数10】

$$P_A = \frac{a}{N} \quad P_B = \frac{b}{N} \quad P_C = \frac{c}{N}$$

【0050】であり、a、b、cの値は知ることができるから、 P_A 、 P_B 、 P_C も得られる。ここで、P'は P_A 、 P_B 、 P_C が独立な確率事象である場合の結合確率、即ち、

$$P' = P_A \times P_B \times P_C$$

で推定することができる。

【0051】代替検索条件Qの指標値S(Q)は、PのP'に対する比であり、

【0052】

【数11】

$$S(Q) = \frac{P}{P'} = \frac{P}{P_A \times P_B \times P_C} = \frac{z N^2}{a b c}$$

【0053】となる。従って、上記指標値S(Q)は既知の値を基に算出できる。上記電話帳検索システムの例では、検索項目は、A、B、Cの3項目であったが、項目数がいくつであっても同様に、

【0054】

【数12】

$$S(Q) = \frac{P}{P'} = \frac{P}{P_A \times P_B \times P_C \times \dots}$$

【0055】によって計算できる。また、上記の議論より代替検索条件の適切さを比較するには、指標値の大小関係を比較すれば充分であるから、代替検索条件Q($Q = A \cap B \cap C \cap \dots$)を満たすデータが存在する確率をPとし、Qの各検索キーA、B、C…のみを満たすデータが存在する確率を P_A 、 P_B 、 P_C 、…とし、代替検索条件Qの評価関数F(Q)を、

【0056】

【数13】

$$F(Q) = \frac{P}{P_A \times P_B \times P_C \times \dots}$$

【0057】とし、任意の代替条件 Q_1 、 Q_2 に対して $F(Q_1) > F(Q_2)$ ならば、 $S(Q_1) > S(Q_2)$ が成り立つ任意の評価関数S(Q)を指標値として用いることができる。

図4は、本発明の一実施例の情報検索システムの構成を示す。同図に示す情報検索システムは、複数項目からなる検索条件が入力される入力部31、入力された検索条件を満たすデータが存在しない場合、検索条件を部分的に緩和することによって構成される代替検索条件を生成する条件生成部32、複数発生する代替検索条件から指標値に基づいて最適な代替検索条件を判断する代替条件判定部33、指標値を算出する指標値算出部34、検索対象データを格納するデータベース35、データベース35を検索する検索部36、検索結果または、検索条件を表示する出力部37、及び各部を制御する制御部38より構成される。

【0058】図5は、本発明の一実施例の情報検索システムの動作を示すフローチャートである。まず、複数の項目からなる検索条件が入力部31より入力される(ステップ41)。ここで例えば、検索条件の項目として、電話帳検索システムには、“住所”、“職業”、“名義”等がある。制御部38が入力条件を満たすデータがデータベース35に存在するかを検索部36に対して指示する。検索部36は、入力された検索条件によりデータベース35を検索し、入力された検索条件を満たすデータが存在している場合は(ステップ42、Yes)、データベース35よりデータを得る。また、検索条件を満たすデータがデータベース35にない場合は(ステップ42、No)、代替検索条件生成部32により、入力された検索条件を部分的に緩和して生成される検索条件を代替検索条件とする(ステップ43)。

【0059】検索部36は、生成された代替検索条件でデータベース35を繰り返し検索する。または、指標値算出部34は、各代替検索条件の指標値を算出し(ステップ44)、最適条件判断部33は、指標値の最も大きい代替検索条件を最適な代替検索条件と判断する(ステップ45)。最適な代替検索条件が判断されると、制御部38を介して、検索部36が再度データベース35を検索し、その検索結果が出力部37により提示される。

【0060】以下、上記電話帳データベース35において、上記入力条件1が入力された場合の処理の流れを具体的に説明する。

入力条件1

住所：千代田区 北の丸公園

職業：美術館

名義：キンダイビジュツカン

データベースのデータ総数 $N = 942837$ 、入力条件1に対する適合解の数 $X = 2$ である。2件の適合解の掲載条件を以下に示す。

掲載条件1-1

住所：千代田区 北の丸公園

職業：美術館

名義：コクリツキンダイビジュツカン

掲載条件 1-2

住所：千代田区 北の丸公園

職業：美術館

名義：コクリツキンダイビジュツカンコウゲイブ

まず、入力条件 1 で電話帳データベース 35 を検索する。

【0061】入力条件 1 を満たす掲載データは存在しないため、(処理フロー“NO”)、代替条件生成部 32 において入力条件の部分的削除による代替検索条件を全て生成する。生成した代替検索条件は図 13 に列挙したものと同様である。次に、検索部 36 は、全ての代替検索条件でデータベースを再検索する。各代替検索条件とその検索解の数を図 6 に示す。また、図 6 の◎を付した代替検索条件は適合解の数が最大 (2 件) で、不適合解の数が最小 (0 件) となる最適な代替検索条件である。図 6 の○を付した代替検索条件は、検索解に適合解を含むが不適合解が最小ではない代替検索条件である。

【0062】上記の結果のうち検索解の数が 0 となる代替検索条件は明らかに不適である。次に、指標値算出部 34 により代替検索条件の適切さの指標値を算出する。まず、入力条件が要求するデータがデータベースに存在しないと仮定した場合に、代替検索条件 Q を満たすデータが存在する確率の推定値 P' を算出する。本実施例において、代替検索条件 Q が「住所：A、職業：B、名義：C」であるとき、即ち、 $Q = A \cap B \cap C$ であるとき、住所 A のみを満たすデータの数 a は、代替検索条件「住所：A、職業：なし、名義：なし」の検索解の数であり、同様に職業：B のみを満たす検索解の数 b は代替検索条件「住所：なし、職業：B、名義：なし」の検索解の数、名義：C のみを満たす検索解の数 c は、代替検索条件「住所：なし、職業：なし、名義：C」の検索解の数である。これらの値は、前過程で得た、他の代替検索条件の検索解の数を参照することによって得られる。

例えば、

代替検索条件 1-0

住所：千代田区

職業：なし

名義：キンダイ

において、「住所：千代田区」のみを満たすデータの数 a は、

代替検索条件 1-1

住所：千代田区

職業：なし

名義：なし

の検索解の数 73549、

「職業：なし」のみを満たすデータの数 b は、

代替検索条件 1-2

住所：なし

職業：なし

名義：なし

10 の検索解の数 942873、

「名義：キンダイ」のみを満たすデータの数 c は、

代替検索条件 1-3

住所：なし

職業：なし

名義：キンダイ

の検索解の数 281 である。

【0063】また、データ総数 N は、データ総数 942873 である。このようにして得た a, b, c, N から上記推定値 P' を算出できる。

20 【0064】

【数 14】

$$P' = P_A \times P_B \times P_C = \frac{a}{N} \times \frac{b}{N} \times \frac{c}{N}$$

【0065】次に、代替検索条件 Q' を満たす検索解が存在する確率 P は、Q の検索解の数を z とすると、

【0066】

【数 15】

$$P = \frac{z}{N}$$

30 【0067】で求めることができる。例えば、代替検索条件 1-0 の検索解の数は 18 であるから、上記代替検索条件 1-0 の検索解が存在する確率 P は、 $z = 18$ より直ちに求められる。本実施例では、指標値 S (Q) として、 P' に対する P の比の対数値を用いる。対数値を用いても各代替検索条件間の指標値の大小関係は変わらない。

【0068】 $S(Q) = \log_2 [P(Q) / P'(Q)]$

指標値算出部 34 では、上記指標値を 1 件以上の検索解を出力する全ての代替検索条件について算出する。入力条件 1 の代替検索条件に上記指標値を算出した結果を表 1、表 2 に示す。

【0069】

【表 1】

40

問い合わせ例 1：検索項目数が 1 つの場合

| 入力条件 | | 検索解の数 | 指標値 |
|------|-----------|-------|-----|
| 名義 | キンダイビジュツ | 2 | 0 |
| | キンダイビジュ | 2 | 0 |
| | キンダイビジ | 3 | 0 |
| | キンダイビ | 19 | 0 |
| | キンダイ | 271 | 0 |
| | キンダ | 281 | 0 |
| | キン | 2397 | 0 |
| | キ | 26212 | 0 |
| 住所 | 千代田区 | 73549 | 0 |
| | 千代田区北の丸公園 | 56 | 0 |
| 職業 | 美術館 | 105 | 0 |

【0070】

【表 2】

問い合わせ例：検索項目が複数の場合

| 入力条件 | | 検索解の数 | 指標値 |
|------------------------|--|-------|-------------|
| 住所：千代田区 名義：キンダイ | | 18 | -0.23 |
| 住所：千代田区 名義：キンダ | | 18 | -0.28 |
| 住所：千代田区 名義：キン | | 283 | 0.59 |
| 住所：千代田区 名義：キ | | 2125 | 0.05 |
| 住所：千代田区北の丸公園 名義：キ | | 1 | -0.63 |
| 職業：美術館 名義：キ | | 1 | -1.54 |
| 住所：千代田区 職業：美術館 | | 13 | 0.66 |
| 住所：千代田区北の丸公園 職業：美術館 | | 2 | <u>8.32</u> |

【0071】計算結果は、上記従来の技術の説明で述べた最適な代替検索条件、

代替検索条件 1

住所：千代田区北の丸公園

職業：美術館

名義：なし

の指標値が最も高い値となっている。

【0072】代替検索条件判定部 33 において、指標値が最も高い上記代替検索条件 1 を最適な代替検索条件と判断する。最後に最適と判断された代替検索条件の検索結果を出力部 37 より出力する。同様に、上記掲載条件 2

「住所：港区 芝

職業：出版社

名義：エヌティティシュッパン」

に対して、上記入力条件 2

「住所：港区 三田

職業：出版社

名義：エヌティティシュッパン」

40 が入力された場合の各代替検索条件とその検索解の数を図 7 に示す。図 7 の◎を付した代替検索条件は、適合解の数が最大（10 件）で不適合解の数が最小（0 件）となる最適な代替検索条件である。図 7 の○を付した代替検索条件は、検索解に適合解を含むが適合解が最小でない代替検索条件である。

【0073】各代替検索条件の指標値の算出結果を表 3、表 4 に示す。

【0074】

【表 3】

| 検索条件 | | 検索解の数 | 指標値 |
|------|-------|-------|-----|
| 名義 | エヌティー | 34 | 0 |
| | エヌティ | 557 | 0 |
| | エヌテ | 570 | 0 |
| | エヌ | 1940 | 0 |
| | エ | 23214 | 0 |
| 住所 | 港区 | 71433 | 0 |
| | 三田 | 2637 | 0 |
| 職業 | 出版社 | 5762 | 0 |

【0075】

【表4】

問い合わせ例2：検索項目数が複数の場合

| 入力条件 | 検索解の数 | 指標値 |
|-------------------------|-------|-------------|
| 住所：港区 名義：エヌティー | 4 | 0.68 |
| 住所：港区 名義：エヌティ | 160 | 1.92 |
| 住所：港区 名義：エヌテ | 164 | 1.92 |
| 住所：港区 名義：エヌ | 483 | 1.71 |
| 住所：港区 名義：エ | 2440 | 0.47 |
| 住所：港区三田 名義：エヌティ | 1 | -0.63 |
| 住所：港区三田 名義：エヌテ | 1 | -0.67 |
| 住所：港区三田 名義：エヌ | 8 | 0.56 |
| 住所：港区三田 名義：エ | 80 | 0.30 |
| 職業：出版社 名義：エヌティ | 10 | 1.55 |
| 職業：出版社 名義：エヌテ | 10 | 1.52 |
| 職業：出版社 名義：エヌ | 25 | 1.07 |
| 職業：出版社 名義：エ | 185 | 0.38 |
| 住所：港区 職業：出版社 | 500 | 0.19 |
| 住所：港区三田 職業：出版社 | 8 | -1.01 |
| 住所：港区 職業：出版社 名義：エヌティ | 10 | <u>5.27</u> |
| 住所：港区 職業：出版社 名義：エヌテ | 10 | 5.24 |
| 住所：港区 職業：出版社 名義：エヌ | 11 | 3.61 |
| 住所：港区 職業：出版社 名義：エ | 25 | 1.21 |

【0076】計算結果は、前述の従来の技術で説明した
最適な代替検索条件、
代替検索条件2
「住所：港区
職業：出版社
名義：エヌティ」

の指標値が最大となり、代替検索条件判断部53において、指標値が最も高い上記代替検索条件2を最適な代替検索条件と判断する。

【0077】最後に、上記により最適と判断された代替検索条件の検索結果を出力部37より出力する。上記の

50 ように最適と判断された上記の代替検索条件の検索解集

合は、図6、図7において◎を付した代替検索条件であり、適合解の数が最大で、且つ不適合解の数が最小（0件）となる最適な代替検索条件である。

【0078】また、これらの代替検索条件で再検索を行う前に、利用者に提示する方法もある。上記のように、代替検索条件の指標値を指標とすることにより、最適な代替検索条件を推定し、利用者の要求するデータを表示することができる。図8は、本発明の第2の実施例の知識獲得システムの構成を示す。知識獲得システムは、上記第1の実施例の情報検索装置を有する。知識獲得システムは、曖昧な入力条件を修正し、適合解を検索する情報検索装置300と、情報検索装置300の情報検索部36によって優先的に出力された掲載データが当初の要求に適合すると利用者が承認したことを確認する適合確認部310と、入力条件と掲載データの対応関係の有効性を評価する有効性評価部320と、信頼性が高いと評価された場合には、掲載データの掲載条件の別名として、その入力条件を登録する別名登録部330と、登録情報を記録する別名記憶部340と、各部の動作を制御する制御部360を具備している。

【0079】図9は、本発明の第2の実施例の知識獲得システムの処理の流れを示すフローチャートである。検索過程により、上記の第1の実施例の情報検索装置300の動作により入力条件が要求する適合解を検索し、この際の入力条件が自動修正されると（ステップ801）、適合性確認過程として検索過程で優先的に出力された掲載データが当初の要求に適合すると利用者が承認したことを確認し（ステップ802）、次に有効性記憶過程として、入力条件と掲載データの対応関係の有効性を評価し（ステップ803）、対応関係が有効と判断された場合には、別名登録過程として掲載データ掲載条件の別名として入力条件を別名記憶部340に登録する（ステップ804）。

【0080】図10は、本発明の第3の実施例の知識獲得システムの構成を示す。同図に示す知識獲得システムは、上記第1の実施例の情報検索装置300を有する。知識獲得システムは、情報検索部を有し、曖昧な入力条件を修正し、適合解を検索する情報検索装置300、情報検索部によって優先的に出力された掲載データが当初の要求に適用すると利用者が承認したことを確認する適合確認部310、修正された入力条件と適合性が確認された掲載データの対応関係の累積頻度と掲載データが検索された累積総数を記憶する累積頻度記憶部350、入力条件と掲載データの対応関係の有効性を評価する有効性評価部320、信頼性が高いと評価された場合には、掲載データの掲載条件の別名としてその入力条件を登録する別名登録部330、登録情報を記憶する別名記憶部340及び各部の動作を制御する制御部360とを具備している。

【0081】図11は、本発明の第3の実施例の知識獲

得システムの処理を示すフローチャートである。検索過程において、情報検索装置300により第1の実施例と同様に入力条件が要求する適合解を検索し（ステップ901）、この際、入力条件が自動修正された場合には、適合確認部310が適合性確認過程において、検索過程で優先的に出力された掲載データが当初の要求に適合すると利用者が承認したことを確認し（ステップ902）、累積頻度登録過程において、累積頻度記憶部350に記憶されている適合性を確認された入力条件と掲載データの対応関係の累積頻度とデータが検索された累積総数を更新し（ステップ903）、次に有効性記憶過程として、有効性評価部320が累積頻度が更新された入力条件と掲載データの対応関係の有効性を評価し（ステップ904）、対応関係が有効と判断された場合には、別名登録過程にて、別名登録部330が掲載データの掲載条件の別名として入力条件を別名記憶部340に登録する（ステップ905）。

【0082】次に入力条件6が入力された時、情報検索装置300での検索過程において、入力条件の自動修正が行われ、代替検索条件6が選択され、検索結果として掲載条件6が出力される。

入力条件6

住所：武蔵野市 緑町

職業：フランス料理

名義：ルパレハットリ

代替検索条件6

住所：武蔵野市 緑町

職業：

名義：ルパレハットリ

掲載条件6

住所：武蔵野市 緑町

職業：レストラン

名義：ルパレハットリ

上記の第3の実施例において、適合性確認過程ステップ902では、以下のように検索結果の適合性を確認する。まず、掲載条件のみを表示し、利用者が上記掲載条件6の電話番号が当初の要求に適合するものと承認すれば、利用者が掲載条件6の電話番号の表示を利用者が指示する。表示指示が行われた掲載データは適合データであると確認される。

【0083】「掲載条件6のデータ→職業：フランス料理」のような、利用者の承認が確認された掲載データを入力条件の対応関係を、以下「別名対応関係」と呼ぶ。次に、累積頻度登録過程（ステップ903）において、上記別名対応関係の累積頻度と、上記掲載データが検索された累積総数を1増やす。本実施例では、掲載データを例えば、「00001」とコード化して登録するものとして、累積頻度登録過程（ステップ903）を図12（A）に示す。別名対応関係が初めて発生した場合には、別名対応関係を新たに登録し、累積頻度を1とす

る。この過程を図12(B)に図示する。

【0084】本実施例の有効性評価部320の有効性評価過程での有効性評価方法は、以下のように定義する。同一の掲載条件に関する別名対応リストの累積頻度の総和をNとし、有効性評価の対象となる別名対応リストの累積頻度をnとする。ここで、次の評価式

$$n \geq 1 + N / 10$$

が成り立つとき、上記別名対応関係は、有効と判断されるものとする。

【0085】例えば、図12(A)に図示したように、登録前に「掲載条件6のデータ→職業：フランス料理」なる別名対応関係が2回あり、掲載データが検索された累積総数は9回であったとする。即ち、 $n = 1$ 、 $N = 9$ であり、この段階では上記別名対応関係は評価式を満たしていない。ここで、新たに上記別名対応関係が登録された場合、 $n = 2$ 、 $N = 10$ となり、評価式を満たす。従って、この段階で、上記別名対応関係は有効と判断され、上記別名対応関係は別名記憶部340に登録される。

【0086】一方、図12(B)に図示した場合、登録後も上記評価式を満たさないため、有効性評価部320は有効とは判断せず、上記別名対応関係は別名記憶部340に登録されない。さらに、本発明の知識獲得システムで得られた知識は、システムの構成部とは別のデータベース検索システムや、各種知識処理にも活用できる。

【0087】なお、本発明は、上記の実施例に限定されことなく、特許請求の範囲内で種々変更・応用が可能である。

【0088】

【発明の効果】従来、利用者が行っていた検索結果を検討し、誤りの部分を予測し、検索条件の修正をいう一連のプロセスのフィードバックを、本発明は、利用者に代わって自動的に行うことが可能である。上述のように、本発明によれば、情報検索装置に入力された条件を満たすデータが存在しない場合、複数の代替検索条件の中から最適代替検索条件を推定し、利用者の要求するデータを最も多く含み、かつ、利用者の要求しないデータを最小にする検索解集合、或いはその検索条件を出力することができる。

【0089】また、本発明の知識獲得システムは、別名登録を自動化することにより、検索サービスの有効効率を向上することができる。また、有効な別名を登録することでより曖昧な問い合わせについても対応することができる。例えば、

入力条件7

住所：港区

職業：美術館

名義：キンダイビジュツカン

は、職業のみしか正しい条件が入力されていないため、上記の実施例の情報検索システムで適合解を検索できな

い。しかし、知識獲得システムを併用して入力条件1の検索結果から「名義：キンダイビジュツカン」が別名登録されていれば、入力条件7に対しても上記の実施例の情報検索システムは適合解を出力できるようになる。

【図面の簡単な説明】

【図1】本発明の情報検索装置の原理構成図である。

【図2】本発明の情報検索方法の原理を説明するための図である。

【図3】本発明の知識獲得システムの原理構成図であ

10 る。

【図4】本発明の一実施例の情報検索装置の構成図である。

【図5】本発明の一実施例の情報検索システムの動作を示すフローチャートである。

【図6】問い合わせ例、問い合わせ例が要求するデータ検索項目及び代替検索条件とその検索解の数を示す図（その1）である。

【図7】問い合わせ例、問い合わせ例が要求するデータ検索項目及び代替検索条件とその検索解の数を示す図（その2）である。

20

【図8】本発明の第2の実施例の知識獲得システムの構成図である。

【図9】本発明の第2の実施例の知識獲得システムの処理の流れを示すフローチャートである。

【図10】本発明の第3の実施例の知識獲得システムの構成図である。

【図11】本発明の第3の実施例の知識獲得システムの処理を示すフローチャートである。

30

【図12】本発明の第3の実施例の累積頻度の登録を説明するための図である。

【図13】従来の検索条件の自動総和の例を示す図である。

【符号の説明】

31 入力部

32 条件生成部

33 最適条件判断部

34 指標値算出部

35 データベース

36 検索部

40

37 出力部

38 制御部

100 代替検索条件生成手段

110 代替検索条件取得手段

112 指標値算出手段

113 条件選択手段

130 提示手段

140 検索条件

150 表示手段

200 データベース

50

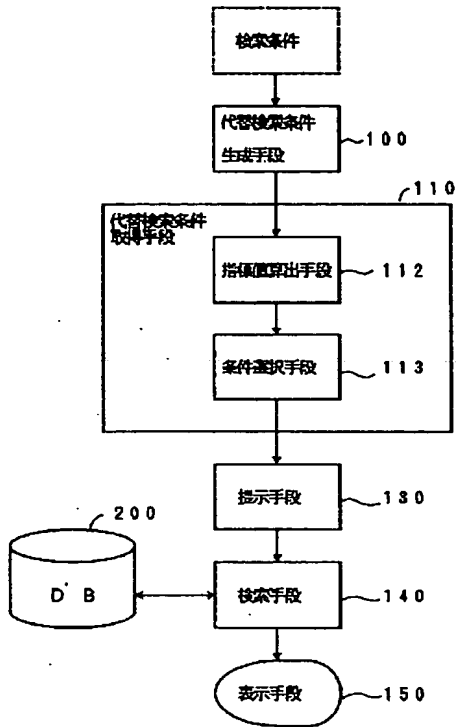
300 情報検索装置

- 310 適合確認手段、適合確認部
- 320 有効性評価手段、有効性評価部
- 330 別名登録手段、別名登録部

- 340 別名記憶部
- 350 累積頻度記憶手段、累積頻度記憶部
- 360 制御部

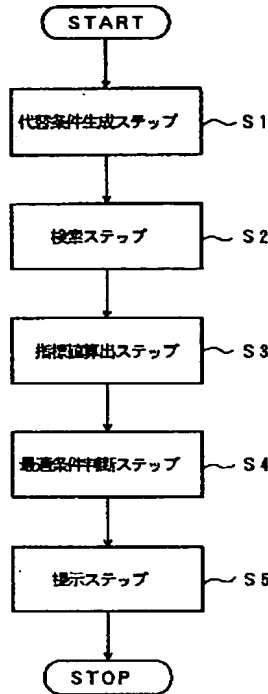
【図1】

本発明の情報検索装置の原理構成図



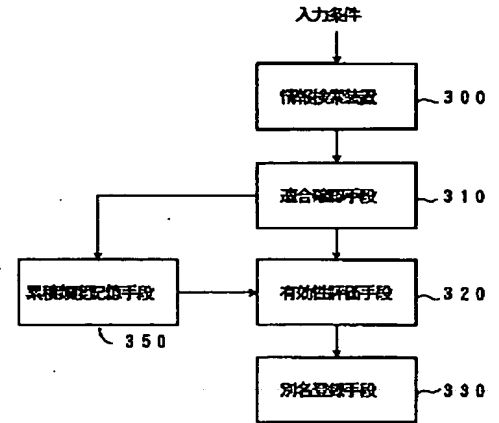
【図2】

本発明の情報検索方法の原理を説明するための図



【図3】

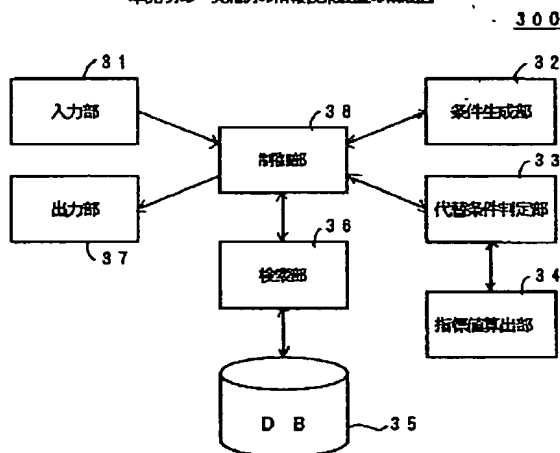
本発明の知識獲得システムの原理構成図



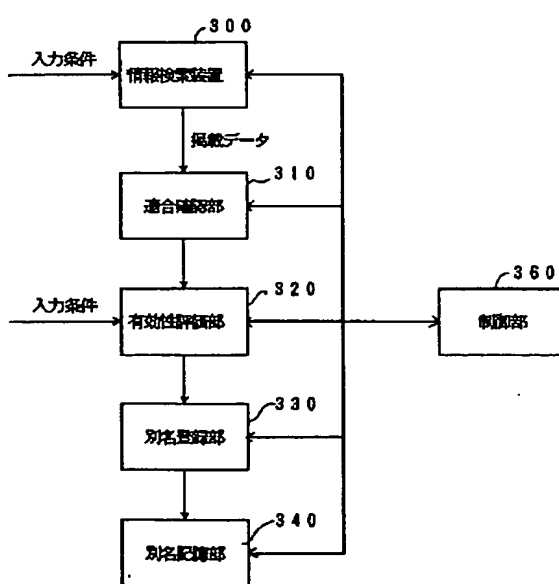
【図4】

【図4】

本発明の一実施例の情報検索装置の構成図

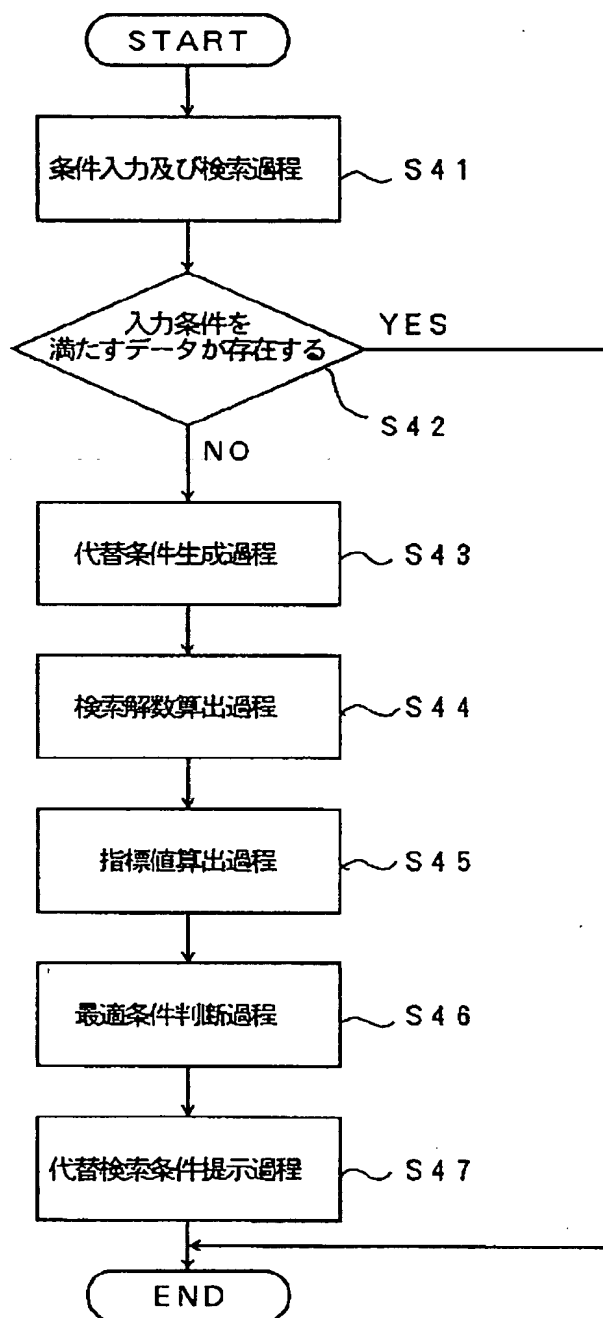


本発明の第2の実施例の知識獲得システムの構成図



【図5】

本発明の一実施例の情報検索システムの動作を示すフローチャート



【図6】

問い合わせ例、問い合わせ例が要求するデータの検索項目
及び代替検索条件とその検索解の数を示す図（その1）

問い合わせ例1

住所：千代田区 北の丸公園
職業：美術館
名義：キンダイビジュツカン

問い合わせ例1が要求するデータの検索項目

住所：千代田区 北の丸公園
職業：美術館
名義：コクリツキンダイビジュツカン

問い合わせ例1の代替検索条件と、検索解の数

| 名義 | 職業 住所 | なし なし | なし 千代田区 | なし 千代田区 北の丸公園 | 美術館 なし | 美術館 千代田区 | 美術館 千代田区 北の丸公園 |
|------------|----------|----------|------------|---------------------|-----------|-------------|----------------------|
| キンダイビジュツカン | | 0 | 0 | 0 | 0 | 0 | 0 |
| キンダイビジュツカ | | 0 | 0 | 0 | 0 | 0 | 0 |
| キンダイビジュツ | | 2 | 0 | 0 | 0 | 0 | 0 |
| キンダイビジュ | | 2 | 0 | 0 | 0 | 0 | 0 |
| キンダイビジ | | 3 | 0 | 0 | 0 | 0 | 0 |
| キンダイビ | | 19 | 0 | 0 | 0 | 0 | 0 |
| キンダイ | | 271 | 18 | 0 | 0 | 0 | 0 |
| キンダ | | 281 | 18 | 0 | 0 | 0 | 0 |
| キン | | 2397 | 283 | 0 | 0 | 0 | 0 |
| キ | | 26212 | 2125 | 1 | 1 | 0 | 0 |
| なし | | ○ 942837 | ○ 73549 | ○ 56 | ○ 105 | ○ 13 | ◎ 2 |

【図 7】

問い合わせ例、問い合わせ例が要求するデータの検索項目
及び代替検索条件とその検索解の数を示す図（その 2）

問い合わせ例 2

住所：港区 三田
職業：出版社
名義：エヌティーティーシュッパン

問い合わせ例 2 が要求するデータの検索項目

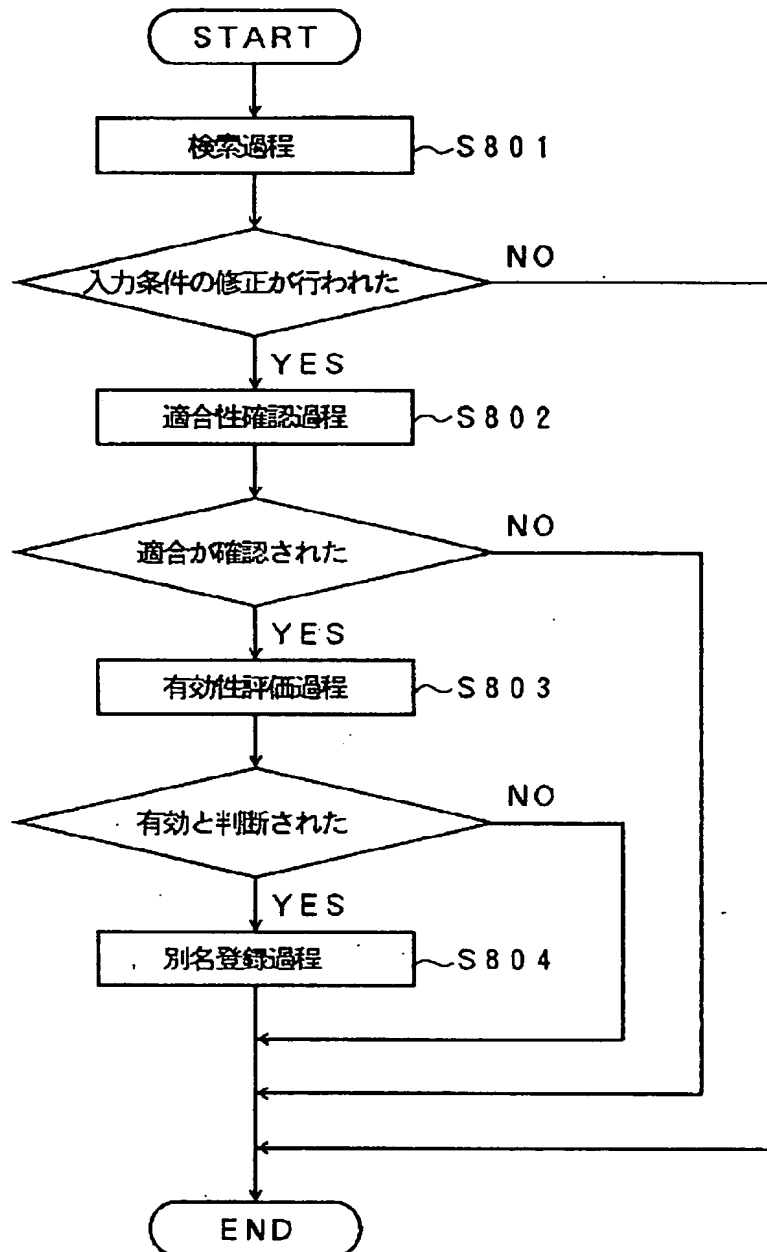
住所：港区 芝
職業：出版社
名義：エヌティティシュッパン

問い合わせ例 2 の代替検索条件と、検索解の数

| 名義 | 職業 住所 | なし なし | なし 港区 | なし 港区 三田 | 出版社 なし | 出版社 港区 | 出版社 港区 三田 |
|---------------|----------|----------|----------|-------------|-----------|-----------|--------------|
| エヌティーティーシュッパン | | 0 | 0 | 0 | 0 | 0 | 0 |
| エヌティーティーシュッパ | | 0 | 0 | 0 | 0 | 0 | 0 |
| エヌティーティーシュッ | | 0 | 0 | 0 | 0 | 0 | 0 |
| エヌティーティーシュ | | 0 | 0 | 0 | 0 | 0 | 0 |
| エヌティーティーシ | | 0 | 0 | 0 | 0 | 0 | 0 |
| エヌティーティー | | 0 | 0 | 0 | 0 | 0 | 0 |
| エヌティーティ | | 0 | 0 | 0 | 0 | 0 | 0 |
| エヌティーテ | | 0 | 0 | 0 | 0 | 0 | 0 |
| エヌティー | | 34 | 4 | 0 | 0 | 0 | 0 |
| エヌティ | ○ | 557 | ○ 160 | 0 | ◎ 10 | ◎ 10 | 0 |
| エヌテ | ○ | 570 | ○ 164 | 1 | ◎ 10 | ◎ 10 | 0 |
| エヌ | ○ | 1940 | ○ 483 | 1 | ○ 25 | ○ 11 | 0 |
| エ | ○ | 23214 | ○ 2440 | 8 | ○ 185 | ○ 25 | 0 |
| なし | ○ | 942837 | ○ 71433 | 80 | ○ 5762 | ○ 500 | 8 |

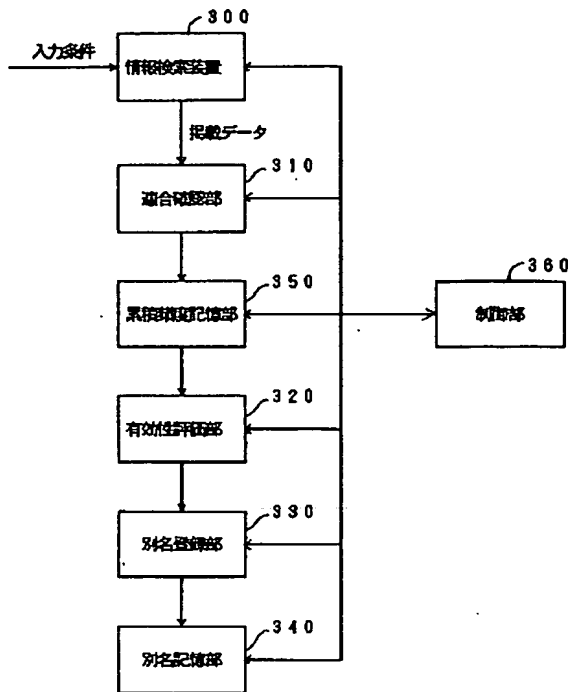
【図9】

本発明の第2の実施例の知識獲得システムの
処理の流れを示すフローチャート



【図10】

本発明の第3の実施例の知識獲得システムの構成図



【図12】

本発明の第3の実施例の累積頻度の登録を説明するための図

「掲載データ(00001)→職業:フランス料理」を登録

既に同様の別名対応関係が登録されている場合

(A)

| | | |
|-----|-----------|----|
| 登録前 | 000001 | 9 |
| | 名義:ハットリ | 1 |
| | 職業:フランス料理 | 1 |
| 登録後 | 000001 | 10 |
| | 名義:ハットリ | 1 |
| | 職業:フランス料理 | 2 |

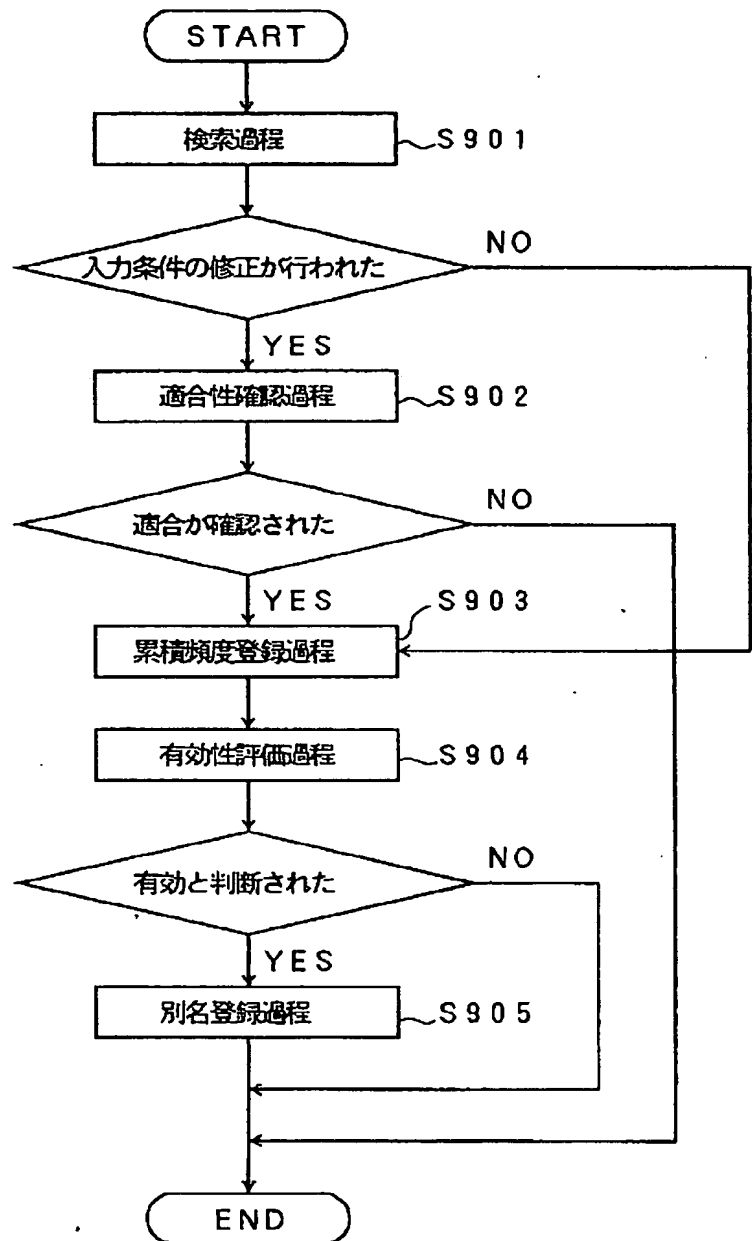
同様の別名対応関係が登録されていない場合

(B)

| | | |
|-----|-----------|----|
| 登録前 | 000001 | 9 |
| | 名義:ハットリ | 1 |
| 登録後 | 000001 | 10 |
| | 名義:ハットリ | 1 |
| | 職業:フランス料理 | 1 |

【図11】

本発明の第3の実施例の知識獲得システムの
処理の流れを示すフローチャート



【図 13】

従来の検索条件の自動総和の例を示す図

問い合わせ例

住所：千代田区北の丸公園
職業：美術館
名義：キンダイビジュツカン

代替条件の候補

| 職業 | 住所 | 名義 | 職業 | 住所 | 名義 |
|-----|-----------|------------|----|-----------|------------|
| 美術館 | 千代田区北の丸公園 | キンダイビジュツカン | なし | 千代田区北の丸公園 | キンダイビジュツカン |
| 美術館 | 千代田区北の丸公園 | キンダイビジュツ | なし | 千代田区北の丸公園 | キンダイビジュツカ |
| 美術館 | 千代田区北の丸公園 | キンダイビジュ | なし | 千代田区北の丸公園 | キンダイビジュツ |
| 美術館 | 千代田区北の丸公園 | キンダイビ | なし | 千代田区北の丸公園 | キンダイビジュ |
| 美術館 | 千代田区北の丸公園 | キンダイ | なし | 千代田区北の丸公園 | キンダイビジ |
| 美術館 | 千代田区北の丸公園 | キンダ | なし | 千代田区北の丸公園 | キンダイビ |
| 美術館 | 千代田区北の丸公園 | キン | なし | 千代田区北の丸公園 | キンダイ |
| 美術館 | 千代田区北の丸公園 | キ | なし | 千代田区北の丸公園 | キンダ |
| 美術館 | 千代田区北の丸公園 | なし | なし | 千代田区北の丸公園 | キン |
| 美術館 | 千代田区 | キンダイビジュツカン | なし | 千代田区北の丸公園 | キ |
| 美術館 | 千代田区 | キンダイビジュツカ | なし | 千代田区北の丸公園 | なし |
| 美術館 | 千代田区 | キンダイビジュツ | なし | 千代田区 | キンダイビジュツカン |
| 美術館 | 千代田区 | キンダイビジュ | なし | 千代田区 | キンダイビジュツカ |
| 美術館 | 千代田区 | キンダイビ | なし | 千代田区 | キンダイビジュツ |
| 美術館 | 千代田区 | キンダイ | なし | 千代田区 | キンダイビジュ |
| 美術館 | 千代田区 | キンダ | なし | 千代田区 | キンダイビジ |
| 美術館 | 千代田区 | キン | なし | 千代田区 | キンダイビ |
| 美術館 | 千代田区 | キ | なし | 千代田区 | キンダイ |
| 美術館 | 千代田区 | なし | なし | 千代田区 | キンダ |
| 美術館 | なし | キンダイビジュツカン | なし | 千代田区 | キン |
| 美術館 | なし | キンダイビジュツカ | なし | 千代田区 | キ |
| 美術館 | なし | キンダイビジュツ | なし | 千代田区 | なし |
| 美術館 | なし | キンダイビジュ | なし | なし | キンダイビジュツカン |
| 美術館 | なし | キンダイビ | なし | なし | キンダイビジュツカ |
| 美術館 | なし | キンダイ | なし | なし | キンダイビジュツ |
| 美術館 | なし | キンダ | なし | なし | キンダイビジュ |
| 美術館 | なし | キン | なし | なし | キンダイビジ |
| 美術館 | なし | キ | なし | なし | キンダイビ |
| 美術館 | なし | なし | なし | なし | キンダイ |
| | | | なし | なし | キンダ |
| | | | なし | なし | キン |
| | | | なし | なし | キ |